

A weighted message-passing algorithm to estimate volume-related properties of random polytopes

Francesc Font-Clos

Centre de Recerca Matemàtica, Edifici C, Campus UAB, E-08193 Bellaterra (Barcelona), Spain

Francesco Alessandro Massucci and Isaac Pérez Castillo

Department of Mathematics, King's College London, WC2R 2LS United Kingdom

In this letter, we introduce a novel message-passing algorithm for a class of problems which can be mathematically understood as estimating volume-related properties of random polytopes. Unlike the usual approach consisting in approximating the real-valued cavity marginal distributions by a few parameters, we propose a weighted message-passing algorithm to deal with the entire function. Various alternatives of how to implement our approach are discussed and numerical results for random polytopes are compared with results using the Hit-and-Run algorithm.

As it is well-known, old and new interesting problems such as Gardner's optimal capacity problem for continuous synaptic couplings [1], von Neumann's problem of linear economics [2], estimation of reaction fluxes from mass-balance equations [3], or reconstruction techniques in compressed sensing [4], can be related, in one way or another, to estimating volume-related properties of random polytopes. In the so-called H-representation, a polytope is defined as the set of points $\mathbf{x} = (x_1, \dots, x_N) \in D \subseteq \mathbb{R}^N$ encapsulated by M hyperplanes $\{(\boldsymbol{\xi}^\mu, \gamma^\mu)\}_{\mu=1}^M$, with $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \dots, \xi_N^\mu)$ the normal vector of the μ -hyperplane. From all possible questions related to polytopes in H-representation, we are particularly interested in that of the volume V and its projection onto each axis relative to V or, in order words, the marginal pdf $P_i(x_i)$. These two quantities can mathematically be written as:

$$V = \int_D d\mathbf{x} \prod_{\mu=1}^M \Theta \left(\sum_{i=1}^N \xi_i^\mu x_i - \gamma^\mu \right), \quad (1)$$

$$P_i(x_i) = \frac{1}{V} \int_D d\mathbf{x}_{\setminus i} \prod_{\mu=1}^M \Theta \left(\sum_{j=1}^N \xi_j^\mu x_j - \gamma^\mu \right), \quad (2)$$

where $\Theta(x)$ is the Heaviside step function and $\mathbf{x}_{\setminus i}$ denotes the vector \mathbf{x} without component i .

Finding efficient methods to obtain reliable estimates for the marginals $P_i(x_i)$ has been, and still is, the main objective of past and present research. In the context of metabolic networks, the area we will focus on to present our work, these methods can roughly be divided into either theoretical approaches of some sort, or Monte Carlo simulations. In the first class of methods, we find Flux Balance Analysis (FBA) [3], which consists in approximating the whole volume V of plausible solutions of the mass-balance equations by a single point, which is selected by optimising an objective function [16]. FBA provides reliable results when applied to elementary micro-organisms in simple milieu conditions (see, for instance, [3, 5]), but the collapsing of V to a single point is too drastic for more complex micro-organisms. To explore

the whole volume V one usually relies on Monte Carlo simulations, as the one introduced in [6], or the Hit-and-Run algorithm [7]. These two algorithms provide a uniform sampling of V , but they are restricted to small networks [17]. For larger networks the MinOver⁺ algorithm has shown to be very promising [8–10], with the sampling becoming more uniform the larger the network is.

Here we want to re-explore the possibility of using message-passing algorithms to estimate $P_i(x_i)$. Message-passing equations, that is cavity equations, were already derived in [8] for the von Neumann problem, and in [11] for mass-balance problem. As the dynamical variables are continuous, it was thought that having a message-passing algorithm for the exact marginals was a daunting task [8], even though fast Fourier transform has been nicely applied in some cases [11].

Let us see how it is possible to introduce a weighted message-passing algorithm for the entire cavity density marginals. Having applications in the area of metabolism in mind, we explain our methodology in the context of the von Neumann model applied to metabolic networks [2, 10, 12]. The application to similar problems is straightforward. In here, a metabolic network is a system of $i = 1, \dots, N$ coupled chemical reactions which produces and consumes $\mu = 1, \dots, M$ metabolites and exchanges them with the environment. Following von Neumann's work [2, 8], as some of the the input will be used to produce output, the ratio of input produced to output consumed cannot be larger than a global growth rate ρ , that is

$$\sum_{i=1}^N (b_i^\mu - \rho a_i^\mu) x_i \geq \gamma^\mu, \quad \mu = 1, \dots, M, \quad (3)$$

where $A = (a_i^\mu)$ and $B = (b_i^\mu)$ are the input and output matrices (e.g. stoichiometric coefficients), x_i is the rate of chemical reaction i , and $\boldsymbol{\gamma} = (\gamma^1, \dots, \gamma^M)$ is the vector of exchanged rates (associated to e.g. transport of metabolites). Given a metabolic network, that is, given a set of stoichiometric coefficients and exchanged rates,

we want to find the volume of solutions V to the set of inequalities (3) at the optimal growth rate. As we do not want to get distracted by unnecessary complications, we denote $\xi_i^\mu \equiv \xi_i^\mu(\rho) = b_i^\mu - \rho a_i^\mu$, ignoring the dependence on ρ and calling $\xi = (\xi_i^\mu)$ the stoichiometric matrix.

In practical situations, the stoichiometric matrix is generally diluted and it is not unreasonable to consider the problem of the volume for diluted random polytopes, meaning that the corresponding bipartite graph associ-

ated to the stoichiometric matrix ξ is tree-like for both i -nodes and μ -nodes. The volume of solutions V is given by (1), with D being the domain region of the reaction rates. This domain is usually determined by the biochemistry of each reaction, that is, reaction rates are usually bounded $x_i \in [x_i^{\min}, x_i^{\max}]$. For sake of simplicity we drop the domain from the notation. Using the cavity method, we obtain the following set of coupled cavity equations [8, 13]:

$$m_\mu^{(i)}(s_i) = \frac{1}{m_\mu^{(i)}} \int ds_{\partial\mu \setminus i} \delta \left(\gamma^\mu - \eta_i^\mu s_i - \sum_{j \in \partial\mu \setminus i} \eta_j^\mu s_j \right) \prod_{j \in \partial\mu \setminus i} P_j^{(\mu)}(s_j), \quad \mu = 1, \dots, M, \quad i \in \partial\mu \quad (4)$$

$$P_i^{(\mu)}(s_i) = \frac{1}{Z_i^{(\mu)}} \prod_{\nu \in \partial i \setminus \mu} m_\nu^{(i)}(s_i), \quad i = 1, \dots, N, \quad \mu \in \partial i \quad (5)$$

where $m_\mu^{(i)}$ and $Z_i^{(\mu)}$ are normalisation constants and the notation e.g. $j \in \partial\mu$ means considering all reactions which participate in the production or consumption of metabolite μ . Here $\mathbf{s} = (x_1, \dots, x_N, z_1, \dots, z_M)$, where the z -variables have been introduced to write the step functions as integrated Dirac deltas, while the matrix η corresponds to the augmented stoichiometric matrix resulting from this transformation. Although this is not a necessary mathematical step, it will be useful for the subsequent discussion.

The actual marginals $P_i(x_i)$ are given by

$$P_i(x_i) = \frac{1}{Z_i} \prod_{\nu \in \partial i} m_\nu^{(i)}(x_i), \quad i = 1, \dots, N, \quad (6)$$

while the meaning of $m_\mu^{(i)}(x_i)$ in eq. (4) is the following: assuming that the marginals $\{P_\ell^{(\mu)}(x_\ell)\}_{\ell \in \partial\mu \setminus i}$ when

reaction i has been removed are known, the probability that, by fixing the value at node i to x_i the inequality μ will be fulfilled, is precisely $m_\mu^{(i)}(x_i)$.

At this point, one would be tempted to use the Dirac delta in eq. (4) in an iteration process with assignments $(\gamma^\mu - \sum_{j \in \partial\mu \setminus i} \eta_j^\mu s_j) / \eta_i^\mu \rightarrow s_i$ as it is usually done in the standard disordered-averaged cavity equations [14]. There are two differences, though: (i) the set of equations (4) is still on an instance; (ii) due to the integration domain, not all possible assignments $(\gamma^\mu - \sum_{j \in \partial\mu \setminus i} \eta_j^\mu s_j) / \eta_i^\mu \rightarrow s_i$ are allowed. Fortunately, the acceptance region in eq. (4) for $m_\mu^{(i)}(s_i)$ is easy to determine as it corresponds to restricting the domain D by the hyperplane μ . Within the acceptance region, we can rewrite equation (4) as follows:

$$m_\mu^{(i)}(s_i) = \frac{1}{m_\mu^{(i)}} \left[\prod_{j=1}^{k_\mu^{(i)}} \int_{\mathcal{R}_j^{(\mu)}(s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu)} ds_{\ell_j} P_{\ell_j}^{(\mu)}(s_{\ell_j} | s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu) \right] \times \delta \left(\gamma^\mu - \eta_i^\mu s_i - \sum_{j=1}^{k_\mu^{(i)}} \eta_j^\mu s_{\ell_j} \right) \prod_{j=1}^{k_\mu^{(i)}} \omega_j^{(\mu)}(s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu) \quad (7)$$

with

$$P_{\ell_j}^{(\mu)}(s_{\ell_j} | s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu) = \frac{P_{\ell_j}^{(\mu)}(s_{\ell_j})}{\omega_j^{(\mu)}(s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu)}, \quad \omega_j^{(\mu)}(s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu) = \int_{\mathcal{R}_j^{(\mu)}(s_{\ell_1}, \dots, s_{\ell_{j-1}}; \gamma^\mu)} ds_{\ell_j} P_{\ell_j}^{(\mu)}(s_{\ell_j}) \quad (8)$$

Here $k_\mu^{(i)} = |\partial\mu \setminus i|$ and $\ell_j \in \partial\mu \setminus i$ for $j = 1, \dots, k_\mu^{(i)}$.

The new expression (7) invites us to propose the following

updating method to solve numerically this equation:

1. Draw s_{ℓ_1} with probability $P_{\ell_1}^{(\mu)}(s_{\ell_1}|\gamma^\mu)$, draw s_{ℓ_2} with probability $P_{\ell_2}^{(\mu)}(s_{\ell_2}|s_{\ell_1};\gamma^\mu)$, ..., draw $s_{\ell_{k_\mu^{(i)}}}$ with probability $P_{\ell_{k_\mu^{(i)}}}^{(\mu)}(s_{\ell_{k_\mu^{(i)}}}|s_{\ell_1},\dots,s_{\ell_{k_\mu^{(i)}}-1};\gamma^\mu)$.
2. Assign $s_i \leftarrow (\gamma^\mu - \sum_{j \in \partial\mu \setminus i} \eta_j^\mu s_j)/\eta_i^\mu$ with weight $\Omega_i^{(\mu)} \equiv \prod_{j=1}^{k_\mu^{(i)}} \omega_j^{(\mu)}(s_{\ell_1},\dots,s_{\ell_{j-1}};\gamma^\mu)$

With this weighted iteration in mind, it is possible to solve numerically the set of equations (4) and (5) directly. We propose two methods: (i) method of histograms. Here one represents $\{P_{\ell_j}^{(\mu)}(s_{\ell_j})\}$ with histograms. Firstly, we use the weighted iteration method in equation (4) to obtain a population of size \mathcal{N} of pairs $\{(s_{i,\alpha}, \Omega_{i,\alpha}^{(\mu)})\}_{\alpha=1}^{\mathcal{N}}$ from which to construct a histogram for $m_\mu^{(i)}(s_i) = (1/\sum_{\alpha=1}^{\mathcal{N}} \Omega_{i,\alpha}^{(\mu)}) \sum_{\alpha=1}^{\mathcal{N}} \Omega_{i,\alpha}^{(\mu)} \delta(s_i - s_{i,\alpha})$. Then use equation (5) to recalculate the histograms of $\{P_{\ell_j}^{(\mu)}(s_{\ell_j})\}$ from the histograms of $\{m_\mu^{(i)}(s_i)\}$; (ii) method of the weighted populations. It is possible to circumvent the step of using and constructing histograms during the iteration process. One simply works directly with a set of equations for the cavity marginals $\{P_{\ell_j}^{(\mu)}(s_{\ell_j})\}$, by combining eqs. (4) and (5). Each of these functions is represented by a weighted population $\{(s_{j,\alpha}, \Gamma_{j,\alpha}^{(\mu)})\}_{\alpha=1}^{\mathcal{N}}$ of size \mathcal{N} . Note that in this case, one Dirac delta appearing in the combined expression gives the assignment for s_j while at the same time, locks the other Dirac deltas, providing an extra weight.

To assess the validity of our approach we have run some numerical checks and compared them with Monte Carlo simulations using the Hit-and-Run algorithm. We entirely focus our numerics in the methods of histograms and postpone a more detailed analysis on the method of weighted populations for another occasion [13]. We have created a random metabolic network of $N = 25$ reactions and $M = 10$ metabolites. The in- and out-degree distributions for reaction and metabolite nodes are chosen to be Poisson with average degree 1.5 and 3.75, respectively. The stoichiometric coefficients ξ_i^μ as well as the exchange fluxes γ^μ are chosen randomly, but ensuring that the volume of solutions is not empty. Finally, we assume that the domain $D = [0, 1]^{25}$.

Results are summarised in fig. 1, which shows all marginals $\{P_i(x_i)\}_{i=1}^{25}$ obtained by the weighted message-passing algorithm, together with the Monte Carlo simulations using the Hit-and-Run algorithm. The comparison is fairly excellent, considering that the cavity equations are based on the Bethe approximation, which does not generally hold for a random graph of finite size, particularly this small.

Looking at the results in fig. 1 we also note that, although originally the domain for each reaction rate is the interval $[0, 1]$, some reaction rates seem to have

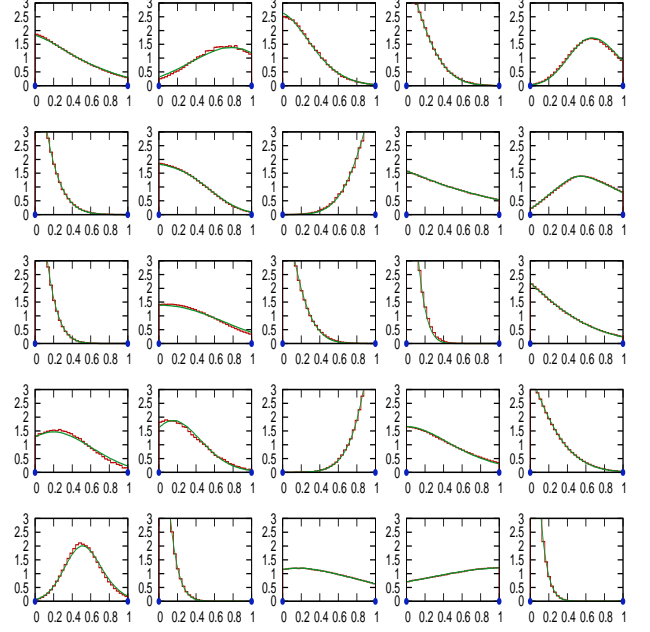


Figure 1: Result for the marginals $P_i(x_i)$ for $i = 1, \dots, 25$ (shown in red and ordered from left to right and up to bottom) using the weighted iteration algorithm within the histogram method for a random metabolic network of $N = 25$ reactions and $M = 10$ metabolites. Our numerical results are compared with Monte Carlo simulations using the Hit-and-Run algorithm (green line). Finally, the solid blue circles are the results of the belief-propagation algorithm for the boundaries of the marginals $\{P_i(x_i)\}$.

a smaller domain region. This can be due because either the polytope further constrains the original domain or rather because there exists a positive but very small probability. It is important to distinguish between what cannot happen and what is unlikely to happen, particularly in metabolism, where a perturbed network can be brought into a region of originally unlikely events (i.e. a disease). To shed some light into this matter, we notice that the set of equations (4) and (5) suggests a way to write a belief-propagation algorithm for the domains $[a_i, b_i] \ni x_i$ of the marginals $P_i(x_i)$. (see [13] for details). The results are also reported in fig. 1 as solid blue circles and indicate that, for this particular instance, the probability in some regions for some variables is very small. To explore further into this matter and to access these regions, we have performed a very simple perturbation of the original metabolic network consisting in changing the domain D by $[0, b]^{25}$ for $b = 1.0, 0.60, 0.30, 0.12$, and $b = 0.10$. Results are shown in fig. 2 for variable 18, where we can see that not only the probability increases, but also that the shape of the pdf changes as well in a non-trivial way.

It is interesting to keep in mind that since a marginal

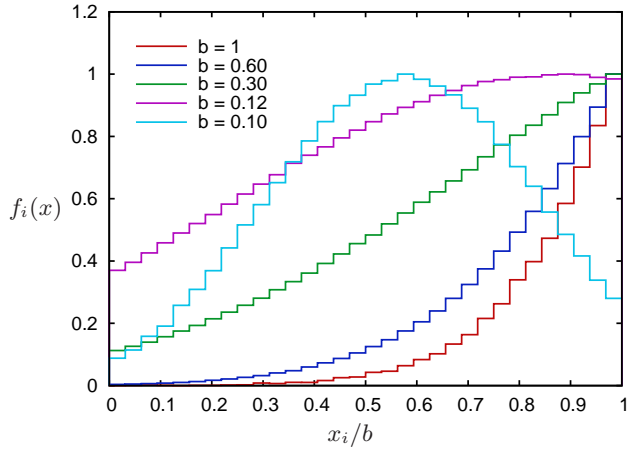


Figure 2: Results for the marginal $P_i(x_i)$ for $i = 18$ and by varying the hypercube $[0, b]^{25}$ for $b = 1.0, 0.60, 0.30, 0.12$, and $b = 0.10$. For a visually easier comparison, the horizontal and vertical axes have been rescaled as x_i/b and $f_i(x) = P_i(x_i)/M_i$ with $M_i = \max_{x \in [0, b]} P_i(x)$, respectively.

$P_i(x_i)$ corresponds to projecting the volume of the polytope onto the x_i -axis, restrictions in the polytope may affect both boundaries $[a, b] \ni x_i$. To follow track of these changes the message-passing equations for the domain [13] are most convenient. As an example, in fig. 3 we have plotted how the domain $[a, b]$ of the marginal $P_i(x_i)$ for variable 22 changes as a in the original domain $[a, 1]^{25}$ of the polytope is varied.

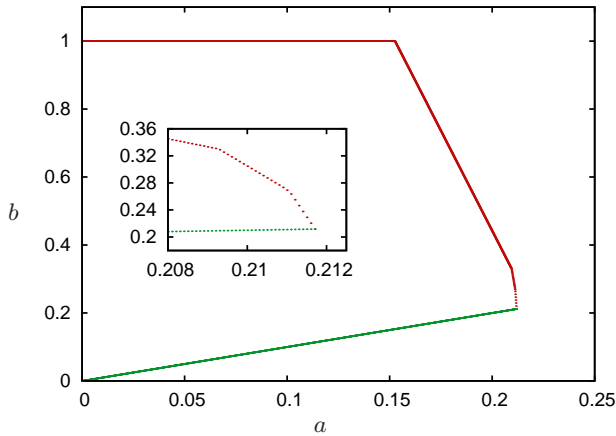


Figure 3: Behaviour of domains $[a, b] \ni x_i$ of marginal $P_i(x_i)$ variable $i = 22$ as the end point a is varied for a perturbed domain $[a, 1]^{25}$. The inset captures a more detailed behaviour of the region where the end points meet.

In this letter we have presented a weighted message-passing algorithm to deal with a class of problems which consists in counting solutions of a set of equations (either equalities or inequalities). The method is designed

to deal with the exact cavity marginals, rather than approximating these with a set of parameters. We have introduced two ways of implementing the method and discussed here one of them, the method of histograms. This method has admittedly the drawback of having to approximate the functions with histograms, but avoids the rejection region characteristic of the problem.

Among the various research lines we are currently considering, we would like to mention two. Firstly, we are currently exploring the implementation of the algorithm by the method of weighted populations [13]. Like in the standard replica symmetric equations in the ensemble, the cavity marginals are naturally represented by a population of random variables. In this way, we avoid having to introduce any discretization of the marginals. Secondly, we are extending this approach to the more interesting case of the double von Neumann problem [15], which is able to incorporate information on the energy balance of the chemical reactions.

-
- [1] E. Gardner and B. Derrida, J. Phys. A **21**, 271 (1988).
 - [2] J. Von Neumann, Ergebn. eines Math. Kolloq. **8**, 73 (1945).
 - [3] K. J. Kauffman, P. Prakash, and J. S. Edwards, Current opinion in biotechnology **14**, 491 (2003).
 - [4] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, Preprint arXiv:1109.4424 (2011).
 - [5] D. Segre, D. Vitkup, and G. M. Church, PNAS **99**, 15112 (2002).
 - [6] N. D. Price, J. Schellenberger, and B. O. Palsson, Biophysical Journal **87**, 2172 (2004).
 - [7] H. Berbee, C. Boender, A. Rinnooy Ran, C. Scheffer, R. Smith, and J. Telgen, Mathematical Programming **37**, 184 (1987).
 - [8] A. De Martino, C. Martelli, R. Monasson, and I. Pérez Castillo, JSTAT **2007**, P05012 (2007).
 - [9] A. De Martino, C. Martelli, and F. A. Massucci, EPL (Europhysics Letters) **85**, 38007 (2009).
 - [10] A. De Martino, D. Granata, E. Marinari, C. Martelli, and V. Van Kerrebroeck, J. Biomed. Biotech. **2010** (2010).
 - [11] A. Braunstein, R. Mulet, and A. Pagnani, BMC bioinformatics **9**, 240 (2008).
 - [12] C. Martelli, A. De Martino, E. Marinari, M. Marsili, and I. Pérez Castillo, PNAS **106**, 2607 (2009).
 - [13] F. Font-Clos, F. Alessandro Massucci, and I. Pérez Castillo, In preparation (2011).
 - [14] M. Mézard and G. Parisi, The European Physical Journal B **20**, 217 (2001).
 - [15] D. De Martino, M. Figliuzzi, A. De Martino, and E. Marinari, Preprint arXiv:1107.2330 (2011).
 - [16] In FBA, one deals with set of equalities instead of inequalities, but the approach is the same. For introduction of FBA see [3].
 - [17] The mixing-time of the Hit-and-Run algorithm goes like $\mathcal{O}(N^3)$, which makes it impractical for large networks